

DOCUMENT RESUME

ED 156 713

TM 007 211

AUTHOR Coffman, William E.
TITLE An Exploratory Study of Group Differences in the Performance of Pupils in Grades 6, 7, 8, and 9 on the Items in the Iowa Tests of Basic Skills.
PUB DATE [Mar 78]
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)
EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; Basic Skills; Black Students; Caucasian Students; Comparative Testing; Elementary Secondary Education; Grade 6; Grade 7; Grade 8; Grade 9; *Group Norms; *Item Analysis; *Racial Differences; Statistical Data; *Test Eas; Test Construction; Test Interpretation; Test Items; Test Results; Test Validity
IDENTIFIERS *Iowa Tests of Basic Skills; *Outliers

ABSTRACT

The Iowa Tests of Basic Skills were administered to over 600 black and white students in grades six through nine, to determine if the test showed bias against minorities. Outliers were identified from test results. Outliers are items which differ from the central core of test items because they fall outside the range expected from a random sampling error. The tests revealed that some items were so difficult for black and white students that the difference in those item scores was essentially random. When these items were removed from the study, there remained 21 items which favored blacks and 16 items which favored whites. However, it is argued that the presence of outliers in distributions of differences in any item statistic does not constitute evidence of item bias, and that the presence of outliers is inevitable whenever groups other than random are compared. A detailed examination of the distributions of differences in item statistics for evidence of clusters which indicate the operation of factors not included in the ability or achievement domain being measured is recommended. In addition, outliers are seen as reflections of educational and cultural influences which should be brought to the attention of teachers and administrators when making educational decisions. (Author/JAC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

William E.
Coffman

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

ABSTRACT

An Exploratory Study of Group Differences in the Performance of
Pupils in Grades 6, 7, 8, and 9 on the Items in the
Iowa Tests of Basic Skills

William E. Coffman

University of Iowa

It is argued that the presence of outliers in distributions of differences in any item statistic does not constitute evidence of item bias, in fact, that the presence of outliers is inevitable whenever groups other than random are compared. What is needed is a detailed examination of the distributions of differences in item statistics for evidence of the presence of clusters that indicate the operation of factors not included in the ability or achievement domain being measured. The argument is supported by data from the responses of black and white pupils at several grade levels to items in the ITBS.

An Exploratory Study of Group Differences in the
Performance of Pupils in Grades 6, 7, 8 and 9
on the Items in the Iowa Tests of Basic Skills

William E. Coffman
University of Iowa

In recent years there has been growing concern about the possibility that educational tests may be biased against various sub-groups of the school population: blacks, Chicanos, pupils from lower socio-economic backgrounds, women--the list can be extended widely. Obviously, the concern is not that different people make different scores on the tests; unless there are relevant differences to be identified, why test at all? The issue, rather, is whether or not the differences identified are differences that have significant diagnostic value in the educational setting, that is, whether or not the differences are valid differences. The concern is that differences may arise as a result of the particular format of certain test questions rather than of the underlying ability or achievement one intends to measure. If, for example, one intends to measure general reading comprehension but the reading passages are from content basically more familiar to or more interesting to one sub-group than to another, their performance on the test may depend on this familiarity rather than on the underlying general reading ability.

A number of criteria have been proposed for identifying the presence of bias in a test. Essentially all rest on the assumption that there is an underlying ability the test is measuring, an ability on which individuals--and perhaps groups--may differ, and that the test development effort has resulted in a core of items that are unbiased. The problem, then, is to look for outliers, that is, for items that differ in some way from the central core.

One way of identifying outliers is to apply some statistical criterion, i.e., the item falls outside the range that would be expected as a result of

PRINTED IN U.S.A.

differences one would expect on the basis of random sampling error. It may be, however, that such procedures are based on an inadequate rationale. In the first place, since tests are designed to identify ability differences--differences that arise as a result of a great variety of differences in educational experiences--it is likely that whenever sub-groups are identified on some basis other than random assignment there will be questions that are outliers in comparison with expectations based on random sampling theory. Furthermore, as a result of sampling error, other items similar to the outliers in what they are measuring will fall within the distribution of sampling error rather than outside. Thus, if one wishes to raise questions regarding bias one needs to do more than identify outliers; one needs also to ask where similar items are falling within the distribution of the item statistic being examined. One needs to ask further, whether the cluster of items thus identified is really evidence of bias or only of performance differences that are just the sorts of differences one is trying to identify by using the test.

Since one cannot ever remove all differences that may arise out of group differences--for example, as a result of the fact that some things are taught more effectively in some schools than in others--and that some differences will inevitably be associated with differences in group experiences not directly relevant to the purpose of the test--for example, that the words girls tend to learn in our culture are sometimes different from the words that boys learn and so if one is not careful, one may build a vocabulary test that is biased against girls--or boys. To avoid bias, then, the test developer will make every effort to provide in his sample of items a balance in the content known to differ for different groups but not to introduce distortions in the

representation of abilities that are the focus of the test development effort.

This paper reports an exploratory study of item bias and balance in the subtests of the Iowa Tests of Basic Skills, levels 12, 13, and 14. The tests were administered to samples of black and white students in a small city. Equal numbers of black and white students were drawn from each of a number of schools to balance for school effects--107 of each in grade 6 (Level 12), 227 in grade 8 (Level 13), 192 in grade 8 (Level 14) and 172 in grade 9 (Level 14). The statistic of concern was the difference between the percentage of the black sample marking the right answer to each question and the percentage of the white sample marking the right answer; however, before computing the difference, the p-values were subjected to the arcsin transformation ($\phi = 2 \arcsin \sqrt{p}$) in order to provide a scale with common sampling characteristics regardless of the particular p-values involved. From statistical theory we reason that if such differences were obtained from random samples from the same population, the differences would be expected to distribute themselves normally around zero with a standard deviation equal to $\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$. Should the differences be generated from random samples from two populations with a difference of K, the distribution of sample differences would be expected to distribute themselves normally around K, again with a standard deviation (S.D.) of $\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$. Should the comparisons of item difficulties be all based on the same two samples, the standard deviation of differences would be expected to be smaller than $\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$ to the extent that there were intercorrelations among the items and should approach zero as the item intercorrelations approached 1.00. Thus to use the theoretical K and S.D. for the sample sizes indicated as a base for comparison with the actual sample K's and S.D.'s observed is to provide a conservative

estimate of the extent to which there is an interaction between items and groups. To the extent that the distribution of differences is symmetrical around the mean difference, we have evidence that balance has been maintained in the test. To the extent that the distribution of differences is skewed, we have evidence of possible bias in favor of the group in the direction of the skew.

Distributions of $\phi_B - \phi_W$ for the items in the eleven subtests of the Iowa Tests of Basic Skills were generated on the University of Iowa computer using a program developed by Scott Olsen, a doctoral student in the Division of Educational Psychology, Measurement, and Statistics. The printout for the grades 6, 7, 8, and 9 samples for the Vocabulary subtest are presented in Tables 1, 2, 3, and 4 respectively. (Similar printouts served as the bases for the summary data for the other ten subtests.) In addition to showing the location of each item in the distribution of arcsin differences, the program computes the mean and standard deviation of the differences and lists out the items that are outliers under four different criteria. For purposes of the present study, the outliers were defined as items falling outside a band plus or minus .5244 standard deviations from the mean (beyond the 80% confidence band.)

For samples of the sizes involved in this study, the theoretical S.D.'s of the differences are .1367 for the 6th grade sample, .0939 for the 7th grade sample, .1021 for the 8th grade sample, and .1078 for the 9th grade sample. These theoretical S.D.'s, the obtained S.D.'s, and the ratios of obtained to theoretical S.D.'s for the 11 subtests of the ITBS are presented in Table 5. With one exception the observed S.D.'s were larger than the

theoretical S.D.'s, suggesting the presence of a group by items interaction. The S.D.'s differ, however, in magnitude. As indicated in Table 6, the largest average ratio was observed for the vocabulary tests and the smallest for the reading and spelling tests. One might argue with some justification that the ordering of the subtests is roughly the same as the order we would obtain if we arranged the subtests from first to last on the basis of the degree to which outside-of-school factors were involved in the development of the ability being measured by the subtests.

The numbers of outliers for the several analyses are presented in Table 7. In general, the distributions are skewed in the direction of the black samples suggesting the possibility of a bias in favor of blacks. However, more careful examination indicates that generally the skew is attributable to the presence in the tests of items that are too difficult for either sample, that is, to items where the differences are attributable primarily to random responses. (These items are identified in the distributions by circles around them.) On the basis of the patterns observed, it seems reasonable to assume that if a lower level of the test had been administered, in most cases the mean differences between blacks and whites would have been even larger, the standard deviations would have been smaller, and the skew, if present at all, would have been in the direction of the white groups.

This does not mean necessarily that the tests are therefore balanced and not biased. To explore the problem further, we examine the outliers to see if we can formulate explanatory hypotheses that might indicate the presence of balance or bias. For example, if outliers in favor of whites appear to be reflecting abilities not directly related to the purpose of the test

while outliers in favor of blacks appear to be directly related to the purpose of the test, we suspect bias; however, if the outliers in both directions appear to require equally relevant abilities, we conclude that the test is balanced but that certain desirable abilities have not yet been developed in members of the black sample to the same extent as in members of the white sample.

To provide an additional frame of reference for interpreting the significance of the patterns observed in the black-white differences, data from grades 7, 8, and 9 for two different junior high schools in a second small city were analyzed. Standard deviations of the arcsin differences and the ratios of observed to theoretical arcsin differences for the eleven ITBS subtests for the three grades are presented in Table 8. Here the ratios range from 1.82 to .86 with a median of 1.17. In contrast, for the grade 7, 8, and 9 black-white samples (Table 5) the ratios range from 1.89 to 1.15 with a median of 1.47. Thus, there appears to be less items by groups interaction for the schools comparisons than for the color comparisons. Furthermore, the order of the ratios is different, as can be seen from Table 9. To the extent that the order reflects differences in curricular emphasis from school to school we would expect the order to change as different pairs of schools are compared.

To the extent that outliers reflect consistent differences in the performances of groups, one would expect them to be replicated across different samples. In Table 10 are presented data regarding replications across the grade 7, 8, and 9 samples for the color comparisons and the school comparisons. The table may be read as follows: Of the vocabulary items in the ITBS involved in the grade 7, 8, and 9 samples used in this study, there were

126 instances where the same item appeared in two of the samples and 39 instances where the same item appeared in three of the samples. For the black-white comparisons there were ten instances where the same item appeared in the extreme 10% of the distribution favoring blacks twice and two instances where the same item appeared three times. There were no replications in the extreme 10% of the distribution favoring whites. For the school comparisons there was one replication in each extreme tail. For the totals across all of the subtests, the comparable figures are 1259 possible pairs, 376 possible triples, 63 pairs and 10 triples in favor of blacks, 22 pairs and 2 triples in favor of whites, 16 pairs in favor of school 1 and 10 pairs in favor of school 2. For random samples, the probability of a pair is .01 and the probability of a triple is .001. Thus, it appears that the number of observed replications for the data from the schools comparison is about what one would expect by chance. On the other hand, the numbers observed for the data from the color comparisons are clearly greater than chance.

As has been noted, to some extent the greater number of replications in the tails of the distributions favoring blacks can be accounted for by the presence in the tests of items so difficult for both samples that differences were essentially random. When these items are removed from consideration, there remain 21 items appearing to favor blacks and 16 appearing to favor whites. This certainly does not indicate a weighting of outliers favoring the white samples. It appears, therefore, that unless the outliers in one direction are judged to be equally appropriate for measuring the underlying achievement for both groups while the outliers in the other direction are judged to be more appropriate for measuring one group than another, then we are dealing with

evidence of balance rather than bias in the tests. The outliers have been identified and can be judged by panels representing the appropriate groups. Frankly, we see the outliers primarily as reflections of educational and cultural influences that should be brought to the attention of teachers and administrators in the schools for use as guides in making educational decisions, not as biased content that should be removed from the tests. The final judgment, however, must be made by the users of the tests.

Table 1

1974-1975 ITBS FALL FORM 05

07/27/77 PAGE 1

TEST V VOCABULARY
 GRADE 06
 LEVEL 12

BLACKS = NB = 107
 WHITES = NW = 107
 $SQR(1/NB+1/NW) = 0.13672$

ITEMS = 46
 MEAN = -0.33583
 STD DEV = 0.18617

INTERVAL FOR DIFFERENCE	COUNT	HISTOGRAM OF ITEMS
BELOW -0.70	0	
-0.70 TO -0.65	1	42
-0.65 TO -0.60	1	47
-0.60 TO -0.55	2	56 65
-0.55 TO -0.50	3	54 59 81
-0.50 TO -0.45	8	43 44 48 52 57 61 73 75
-0.45 TO -0.40	4	45 63 66 68
-0.40 TO -0.35	3	50 69 83
-0.35 TO -0.30	7	51 55 62 71 74 76 80
-0.30 TO -0.25	6	40 41 46 49 79 82
-0.25 TO -0.20	2	60 72
-0.20 TO -0.15	2	64 67
-0.15 TO -0.10	2	70 78
-0.10 TO -0.05	3	53 77 84
-0.05 TO 0.00	0	
0.00 TO 0.05	0	
0.05 TO 0.10	0	
0.10 TO 0.15	0	
0.15 TO 0.20	1	(85)
0.20 & ABOVE	1	(58)

CONFIDENCE INTERVAL ITEMS IN FAVOR OF BLACKS

95% -0.0231 TO -0.7007 58 85
 90% -0.0305 TO -0.6412 77 84
 80% -0.0975 TO -0.5741 77 84
 50% -0.2104 TO -0.4713 53 44 67 70 78

ITEMS IN FAVOR OF WHITES

42
 47 56 65
 48 54 57 59 61 75 81

Table 2

1974-1975 ITBS

FOLL

FORM 05

07/27/77

PAGE 1

TEST V VOCABULARY

GRADE 07

LEVEL 13

BLACKS = NB = 227
 WHITES = NW = 227
 SQR (1/NB+1/NW) = 0.09386

ITEMS = 48
 MEAN = -0.24221
 STD DEV = 0.15846

INTERVAL FOR DIFFERENCE	COUNT	HISTOGRAM OF ITEMS
BELOW -.70	0	
-.70 TO -.65	0	
-.65 TO -.60	0	
-.60 TO -.55	0	
-.55 TO -.50	0	
-.50 TO -.45	5	59 60 62 86 98
-.45 TO -.40	3	74 79 91
-.40 TO -.35	4	67 68 92 103
-.35 TO -.30	5	69 71 79 88 96
-.30 TO -.25	11	63 64 65 66 73 75 77 87 91 93 97
-.25 TO -.20	2	101 105
-.20 TO -.15	3	72 76 80 92 104
-.15 TO -.10	5	61 83 89 95 100
-.10 TO -.05	2	90 102
-.05 TO 0.00	3	58 70 99
0.00 TO 0.05	1	85
0.05 TO 0.10	0	
0.10 TO 0.15	1	84
0.15 TO 0.20	0	
0.20 & ABOVE	1	94

CONFIDENCE INTERVAL

ITEMS IN FAVOR OF BLACKS

ITEMS IN FAVOR OF WHITES

95% = 0.0684 TO -0.5528
 90% = 0.0177 TO -0.5021
 80% = -0.0394 TO -0.4450
 50% = -0.1354 TO -0.3400

84 94
 85
 58 70 99
 61 89 90 100 102

12

59 60 62 86 98
 67 68 74 78 81 92 103

Table 3

1974-1975 ITBS FALL FORM 05

07/27/77 PAGE 1

TEST V VOCABULARY
 GRADE 08
 LEVEL 14

BLACKS = NB = 192
 WHITES = NW = 192
 SQR(1/NB+1/NW) = 0.10206

ITEMS = 48
 MEAN = -0.25685
 STD DEV = 0.19067

INTERVAL FOR DIFFERENCE	COUNT	HISTOGRAM OF ITEMS
BELOW -.70	0	
-.70 TO -.65	0	
-.65 TO -.60	0	
-.60 TO -.55	0	
-.55 TO -.50	3	69 74 81
-.50 TO -.45	5	71 73 75 103 104
-.45 TO -.40	5	92 93 97 98 107
-.40 TO -.35	7	68 77 79 80 86 96 112
-.35 TO -.30	4	76 87 88 91
-.30 TO -.25	2	78 109
-.25 TO -.20	4	82 100 106 (114)
-.20 TO -.15	5	67 72 95 101 113
-.15 TO -.10	2	83 105
-.10 TO -.05	4	(70) (90) 99 (108)
-.05 TO 0.00	1	110
0.00 TO 0.05	2	84 102
0.05 TO 0.10	3	(85) 89 (111)
0.10 TO 0.15	0	
0.15 TO 0.20	0	
0.20 & ABOVE	1	94

CONFIDENCE INTERVAL

ITEMS IN FAVOR OF BLACKS

ITEMS IN FAVOR OF WHITES

95% 0.0000 TO -0.6306
 90% 0.0558 TO -0.5626
 80% -0.0123 TO -0.5007
 50% -0.1263 TO -0.3854

94 111
 89
 84 85 102
 70 83 90 99 108 110

69 74 81
 68 71 73 75 77 92 93 97 98 103 104 107 112

Table 4

1974-1975 ITBS FALL FORM 05 07/27/77 PAGE 1

TEST V VOCABULARY
GRADE 09
LEVEL 14BLACKS = NB = 172
WHITES = NW = 172
SOR (1/NB+1/NW) = 0.10783ITEMS = 48
MEAN = -0.29621
STD DEV = 0.19400

INTERVAL FOR DIFFERENCE	COUNT	HISTOGRAM OF ITEMS
BELOW -.70	0	
-.70 TO -.65	0	
-.65 TO -.60	0	
-.60 TO -.55	2	68 93
-.55 TO -.50	5	67 81 87 89 114
-.50 TO -.45	4	71 80 92 95
-.45 TO -.40	8	65 79 86 96 102 103 104 105
-.40 TO -.35	4	73 75 83 91
-.35 TO -.30	1	93
-.30 TO -.25	6	74 78 90 97 106 113
-.25 TO -.20	5	70 76 77 88 112
-.20 TO -.15	3	100 101 107
-.15 TO -.10	3	82 85 108
-.10 TO -.05	1	99
-.05 TO 0.00	0	
0.00 TO 0.05	2	94 110
0.05 TO 0.10	3	72 84 111
0.10 TO 0.15	0	
0.15 TO 0.20	1	109
0.20 & ABOVE	0	

CONFIDENCE INTERVAL

95% = 0.0000 TO -0.6764
90% = 0.0720 TO -0.6144
80% = -0.0479 TO -0.5445
50% = -0.1854 TO -0.4270

ITEMS IN FAVOR OF BLACKS

72 109
84 110 111
94
82 85 99 100 101 108

ITEMS IN FAVOR OF WHITES

68 93
67 71 79 80 81 87 89 92 95 102 105 114

Table 5

Standard Deviations of $\phi_B - \phi_W$ and Ratios of Observed
to Theoretical Standard Deviations (in parentheses) for
Samples of Pupils in Grades 6, 7, 8, and 9 for Items in ITBS Subtests

	Grade 6 S.D. Ratio	Grade 7 S.D. Ratio	Grade 8 S.D. Ratio	Grade 9 S.D. Ratio
V	.188 (1.37)	.160 (1.70)	.193 (1.89)	.196 (1.81)
R	.174 (1.27)	.108 (1.15)	.138 (1.35)	.146 (1.35)
L-1	.154 (1.12)	.123 (1.31)	.139 (1.36)	.142 (1.31)
L-2	.145 (1.06)	.143 (1.52)	.176 (1.73)	.163 (1.56)
L-3	.136 (.99)	.149 (1.59)	.165 (1.62)	.178 (1.65)
L-4	.193 (1.41)	.125 (1.33)	.192 (1.88)	.189 (1.75)
W-1	.204 (1.49)	.159 (1.69)	.161 (1.58)	.154 (1.43)
W-2	.194 (1.42)	.131 (1.39)	.169 (1.66)	.183 (1.69)
W-3	.167 (1.22)	.123 (1.31)	.147 (1.44)	.152 (1.41)
M-1	.161 (1.18)	.128 (1.36)	.147 (1.44)	.128 (1.19)
M-2	.177 (1.29)	.138 (1.47)	.185 (1.81)	.150 (1.39)
Sample Size	107	227	192	172
Theoreti- cal S.D.	.137	.094	.102	.108

Table 6

Averages of the Ratios of Standard Deviations of $\phi_B - \phi_W$ for
Grade 6, 7, 8, and 9 Samples to Theoretical Standard Deviations
Assuming Random Sampling

<u>Subtest</u>		<u>Average Ratio</u>
Vocabulary	(V)	1.69
Usage	(L-4)	1.59
Maps	(W-1)	1.55
Graphs and Tables	(W-2)	1.54
Arithmetic Problems	(M-2)	1.49
Punctuation	(L-3)	1.47
Capitalization	(L-2)	1.46
Reference Materials	(W-3)	1.35
Math Concepts	(M-1)	1.29
Reading	(R)	1.28
Spelling	(L-1)	1.28

Table 7

Numbers of Outliers (80% confidence band) in Distributions of $\phi_B + \phi_W$
for Items in the ITBS for Samples in Grades 6, 7, 8, and 9

(Numbers in parentheses are numbers after eliminating items difficult for both samples)

TEST	Grade								Total	
	6		7		8		9		B	W
	B	W	B	W	B	W	B	W		
V	4 (2)	4	6 (5)	5	6 (4)	3	6 (5)	2	22 (16)	14
R	6 (5)	6	9 (9)	9	10 (9)	7	10 (9)	7	35 (32)	29
L-1	3 (2)	5	6 (4)	3	6 (5)	5	6 (4)	5	21 (15)	18
L-2	6 (6)	5	7 (6)	5	6 (4)	3	4 (4)	6	23 (20)	19
L-3	4 (3)	5	4 (1)	4	6 (3)	4	3 (2)	5	17 (9)	18
L-4	1 (0)	1	3 (1)	3	2 (2)	3	2 (0)	4	8 (3)	11
W-1	4 (2)	6	5 (2)	2	4 (1)	6	4 (3)	4	17 (8)	18
W-2	4 (1)	2	4 (3)	3	4 (1)	2	3 (1)	0	15 (6)	7
W-3	6 (2)	7	5 (3)	4	6 (3)	6	5 (4)	4	22 (12)	21
M-1	5 (2)	4	6 (1)	4	5 (2)	4	6 (4)	4	22 (9)	16
M-2	5 (2)	2	4 (3)	4	5 (4)	2	4 (4)	3	18 (13)	11
Total Outliers	48 (27)	47	59 (29)	46	60 (36)	45	53 (32)	44	220 (124)	182
Percentage	9.9 (5.5)	9.7	11.8 (5.8)	9.2	11.8 (7.1)	8.9	10.5 (6.3)	8.7	11.0 (6.2)	9.1
No. of Items	487		500		507		507		2001	

Table 8

Standard Deviations of $\phi_1 - \phi_2$ and Ratios of Observed to
Theoretical Standard Deviations (in parentheses) for Samples of
Pupils in Grades 7, 8, and 9 in Two Schools for Items in ITBS Subtests

	Grade 7	Grade 8	Grade 9
	S. D. Ratio	S. D. Ratio	S. D. Ratio
V	.140 (1.16)	.174 (1.17)	.202 (1.17)
R	.147 (1.23)	.180 (1.21)	.205 (1.19)
L-1	.181 (1.51)	.155 (1.05)	.162 (.94)
L-2	.147 (1.22)	.162 (1.10)	.260 (1.50)
L-3	.131 (1.09)	.213 (1.44)	.195 (1.13)
L-4	.103 (0.86)	.149 (1.00)	.160 (0.93)
W-1	.168 (1.40)	.176 (1.19)	.175 (1.01)
W-2	.119 (0.99)	.207 (1.39)	.204 (1.18)
W-3	.122 (1.02)	.138 (0.93)	.188 (1.09)
M-1	.187 (1.56)	.219 (1.82)	.222 (1.28)
M-2	.154 (1.28)	.151 (1.01)	.159 (0.92)
N (School 1)	136	88	49
N (School 2)	142	94	105
Theoretical	.120	.148	.173

Table 9

Averages of the Ratios of Standard Deviations of $\phi_1 - \phi_2$ to
Theoretical Standard Deviations for Grade 7, 8, and 9
Samples from Two Schools for Items in ITBS Subtests

Math Concepts	(M-1)	1.55
Capitalization	(L-2)	1.27
Punctuation	(L-3)	1.22
Reading	(R)	1.21
Maps	(W-1)	1.20
Charts and Graphs	(W-2)	1.19
Vocabulary	(V)	1.17
Spelling	(L-1)	1.17
Arithmetic Problems	(M-2)	1.07
Reference Materials	(W-3)	1.01
Usage	(L-4)	.93

Table 10

Number of Items Falling in the Extreme 10% of the Distributions
of Arcsin Differences Twice or Three Times (in parentheses), and Numbers
of Possible Pairs and Possible Triples for Grade 7, 8, and 9 Samples of Blacks
and Whites and Samples of Two Different Schools for Subtests of the ITBS

<u>ITBS Subtest</u>	<u>Possible Pairs</u>	<u>Possible Triples</u>	<u>Samples Black - White</u>		<u>Sample School 1 - School 2</u>	
V	126	39	10 (2)	0	1	1
R	198	59	7 (1)	2	3	4
L-1	126	39	8 (1)	2	2	0
L-2	110	33	9 (2)	1	0	0
L-3	110	33	4	2	2	1
L-4	74	21	0	5 (1)	1	1
W-1	104	31	9 (2)	1	1	0
W-2	72	22	6 (1)	2	1	0
W-3	145	43	6 (1)	5 (1)	2	2
M-1	112	32	2	2	1	0
M-2	82	24	2	0	2	1
Totals	1,259	376	63 (10)	22 (2)	16	10
In 6th Grade				7		